



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2014

**Texte schreiben 2014 – Standardisierte Erfassung von Schreibkompetenzen.
Bericht zuhanden des Departements Bildung, Kultur und Sport des Kantons
Aargau**

Oostlander, J ; Wespi, Barbara

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-124573>

Scientific Publication in Electronic Form

Published Version

Originally published at:

Oostlander, J; Wespi, Barbara (2014). Texte schreiben 2014 – Standardisierte Erfassung von Schreibkompetenzen. Bericht zuhanden des Departements Bildung, Kultur und Sport des Kantons Aargau. Zürich: Institut für Bildungsevaluation.



**Universität
Zürich** ^{UZH}

**Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich**

Texte schreiben 2014 – Standardisierte Erfassung von Schreibkompetenzen

Bericht zuhanden des Departements Bildung, Kultur und Sport
des Kantons Aargau

Jeannette Oostlander & Barbara Wespi

Zürich, 4. Juli 2014

Institut für Bildungsevaluation
Assoziiertes Institut der Universität Zürich
Wilfriedstrasse 15
8032 Zürich

Tel: 043 268 39 60
Fax: 043 268 39 67
www.ibe.uzh.ch

Barbara.Wespi@ibe.uzh.ch
Jeannette.Oostlander@ibe.uzh.ch

Inhalt

| | | |
|----------|--|-----------|
| 1 | Ausgangslage | 4 |
| 2 | Durchführung | 4 |
| 2.1 | Schreibauftrag und Themen | 5 |
| 3 | Beurteilung der Texte | 7 |
| 3.1 | Beurteilungsraster | 7 |
| 3.2 | Beurteilungskriterien | 8 |
| 4 | Qualität des Beurteilungsrasters | 10 |
| 4.1 | Analyse der Beurteilungskriterien | 10 |
| 4.2 | Eindimensionalität des Beurteilungsrasters | 13 |
| 5 | Beurteilungszuverlässigkeit | 13 |
| 5.1 | Zusammensetzung des Korrekturteams | 13 |
| 5.2 | Verfahren der Mehrfachkorrekturen | 14 |
| 5.3 | Übereinstimmung der Bewertung | 14 |
| 5.4 | Strenge der Bewertungen | 15 |
| 5.5 | Auswahl und Schwierigkeit der Themen | 16 |
| 6 | Testergebnisse | 18 |
| 6.1 | Verteilung der Ergebnisse auf der Stellwerkskala | 18 |
| 6.2 | Verteilung der Ergebnisse nach Geschlecht | 19 |
| 6.3 | Ergebnisse nach Erstsprache | 19 |
| 6.4 | Ergebnisse nach Schultyp | 20 |
| 7 | Beispieltexte zu den einzelnen Kompetenzniveaus | 21 |
| 8 | Fazit | 33 |
| 9 | Glossar: Statistische Kennwerte | 35 |

1 Ausgangslage

Im Kanton Aargau wird das Testsystem «Stellwerk» im Rahmen von Check 8 zur schultypenunabhängigen Leistungsbeurteilung eingesetzt. Stellwerk umfasst Tests für die Fachbereiche Deutsch, Englisch, Französisch, Mathematik sowie Natur und Technik. Die Stellwerk-Tests werden ausschliesslich am Computer gelöst, wobei es sich um adaptive Tests handelt, welche sich an die Fähigkeiten der Schülerinnen und Schüler anpassen. Adaptive Tests haben den Vorteil gegenüber dem traditionellen «Papier-Bleistift-Test», dass die Objektivität der Durchführung gesichert ist und dass der Computer bei der Korrektur keine Fehler macht. Computergestützte Tests haben allerdings insofern Nachteile, als dass sie nur reproduktive Fähigkeiten testen.

Produktive Fähigkeiten – wie ausführliche Antworten auf offene Fragen oder ganze Texte – können am Computer bis jetzt nicht zufriedenstellend korrigiert und bewertet werden. Produktive Fähigkeiten können deshalb am Computer nicht getestet werden. Aus diesem Grund werden produktive Fähigkeiten nach wie vor als «Papier-Bleistift-Test» durchgeführt. Damit eine möglichst standardisierte Auswertung des Stellwerk-Modul «Texte schreiben» gewährleistet werden kann, werden die Texte der Schülerinnen und Schüler durch das Institut für Bildungsevaluation, assoziiertes Institut der Universität Zürich, bewertet. Somit findet eine standardisierte Erfassung der sprachlichen Kompetenzen im Lehrplanbereich «Texte schreiben» statt.

2 Durchführung

Die Durchführung des Moduls «Texte schreiben» fand im Kanton Aargau am 28. und 29. April 2014 statt. Die Teilnahme war für alle 8. Klassen freiwillig. Den Schülerinnen und Schülern wurden drei Themen vorgelegt, wovon sie ein Thema auswählen konnten. Dazu schrieben sie während 60 Minuten einen Text, der in der Regel eine bis drei Seiten umfasste. Insgesamt verfassten 4377 Schülerinnen und Schüler einen Text. Die Texte wurden von den Lehrpersonen kopiert und die Originale dem Institut für Bildungsevaluation zur Beurteilung zugestellt.

Die Texte der Schülerinnen und Schüler wurden am Institut für Bildungsevaluation von sieben eigens dafür geschulten Personen – Lehrpersonen, Germanistinnen und Germanisten – anhand eines standardisierten Beurteilungsrasters bewertet. Die erreichte Punktzahl im Modul «Texte schreiben» wurde in die Ergebnisse von Stellwerk integriert, welche von den Lehrpersonen am Computer eingesehen und ausgedruckt werden können. Zugleich wurden die beurteilten Texte an die Lehrpersonen und ihre Schülerinnen und Schüler zurückgesandt. Das ausgefüllte Beurteilungsraster wurde an jeden Aufsatz angeheftet, sodass die Punktevergabe für die Lehrperson und die Schülerinnen und Schüler bei jedem Kriterium einsehbar ist.

2.1 Schreibauftrag und Themen

Mit dem Modul «Texte schreiben» wird angestrebt, die Fähigkeit zu erfassen, Texte verständlich zu formulieren und je nach Zielsetzung adressatengerecht zu schreiben, präzise zu formulieren, überzeugend zu argumentieren oder Sprache ästhetisch ansprechend und kreativ einzusetzen¹.

Nicht jeder Schreibauftrag eignet sich gleich gut für eine standardisierte Erfassung der Schreibkompetenzen von Schülerinnen und Schülern. Je offener ein Auftrag formuliert wird, desto schwieriger gestaltet sich die standardisierte Beurteilung. Aus diesem Grund werden für das Modul «Texte schreiben» Schreibaufträge bevorzugt, welche einen Lebens- beziehungsweise Alltagsbezug haben und mit klaren Aufgaben verbunden sind. Zur Wahl standen drei Themen:

- «Angst»
- «Was soll man anziehen?»
- «Bussen für das Wegwerfen von Abfall?»

Die Themen wurden den Schülerinnen und Schülern mittels einer kurzen Einleitung und drei zu beantwortenden Fragen vorgestellt. Die Fragen dienten den Schülerinnen und Schülern als Leitfaden und halfen bei der Strukturierung der Texte. Die schriftliche Anleitung für die drei Themen ist in den Abbildungen 1 bis 3 dargestellt.

Abbildung 1: Schreibauftrag «Angst»

1. Angst

Du kennst sicher eine Situation, in der du Angst hattest. Erzähle, wie es dazu kam und wie du die Situation bewältigen konntest. Bekamst du Hilfe?

Verfasse eine Geschichte, in der du folgende Fragen beantwortest:

- Was geschah als du Angst hattest?
- Warum hattest du Angst?
- Wie konntest du die Situation bewältigen?

Achte darauf, dass du die Kriterien erfüllst, nach denen der Text bewertet wird.

¹ Harsch, C., Neumann, A., Lehmann, R. & Schröder, K. (2007). Schreibfähigkeit. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 42–62). Weinheim: Beltz

Abbildung 2: Schreibauftrag «Was soll man anziehen?»

2. Was soll man anziehen?

Viele Jugendliche tragen gerne Markenkleider oder Kleider, die Stars, Models oder andere Vorbilder tragen. Einige Jugendliche oder deren Eltern geben sehr viel Geld für Kleider, Schmuck und Schuhe aus. Anderen Jugendlichen ist es nicht so wichtig, wie sie sich kleiden.

Verfasse einen Text dazu, in dem du folgende Fragen beantwortest:

- Was ist dir bei deinem Outfit wichtig und worauf achtest du beim Kauf?
- Welche Beobachtungen machst du bei anderen?
- Was soll man anziehen, was nicht?

Achte darauf, dass du die Kriterien erfüllst, nach denen der Text bewertet wird.

Abbildung 3: Schreibauftrag «Bussen für das Wegwerfen von Abfall»

3. Bussen für das Wegwerfen von Abfall?

Immer mehr Menschen lassen ihren Abfall im öffentlichen Raum liegen: auf der Wiese, im Zug, auf der Strasse... Die Kosten für die Reinigung steigen deshalb stetig an. Sollen in Zukunft Bussen für das Wegwerfen von Abfall im öffentlichen Raum verteilt werden?

Verfasse einen Text dazu, in dem du folgende Fragen beantwortest:

- Warum wird immer mehr Abfall im öffentlichen Raum liegen gelassen?
- Welche Beobachtungen und Erfahrungen machst du zu diesem Thema?
- Was ist deine Meinung dazu: Soll es Bussen geben oder gibt es andere Lösungen?

Achte darauf, dass du die Kriterien erfüllst, nach denen der Text bewertet wird.

Ebenfalls zum Schreibauftrag gehören Angaben darüber, wie die Texte bewertet werden (siehe Punkt 4 in Abbildung 4). Die ersten beiden inhaltlichen Kriterien beziehen sich auf den expliziten Schreibauftrag: Zum einen muss der kommunikative Auftrag erfüllt werden, zum anderen sollte der Text nach den expliziten Vorgaben strukturiert werden (die Texte müssen eine Einleitung, einen Hauptteil und einen Schluss haben). Darüber hinaus wird den Schülerinnen und Schülern auch mitgeteilt, dass die Kriterien Sprachrichtigkeit («Ist dein Text sprachlich richtig?») und Sprachangemessenheit («Ist die Wortwahl passend und abwechslungsreich?») beurteilt werden.

Abbildung 4: Anleitung zum Schreibauftrag

1. Lies die Informationen zu den drei Themen genau durch.
2. Wähle eines der drei Themen aus und verfasse dazu einen Text.
3. Gehe wie folgt vor:
 - Du darfst Notizpapier zur Vorbereitung gebrauchen.
 - Du darfst zuerst einen Entwurf oder ein Mindmap auf dem Notizpapier erstellen.
 - Schreibe deinen fertigen Text gut lesbar auf die ausgeteilten Blätter.
 - Du darfst den Duden oder ein anderes Wörterbuch benutzen.
 - Du hast insgesamt 60 Minuten Zeit.
4. Dein Text wird nach folgenden Kriterien bewertet:
 - Werden in deinem Text alle drei Fragen ausführlich beantwortet?
 - Hat dein Text eine Einleitung, einen Hauptteil und einen Schluss?
 - Ist dein Text klar und verständlich?
 - Ist dein Text sprachlich richtig?
 - Ist die Wortwahl passend und abwechslungsreich?

3 Beurteilung der Texte

3.1 Beurteilungsraster

Zur Beurteilung der Texte im Rahmen des Moduls «Texte schreiben» wurde ein Beurteilungsraster basierend auf den vier Basisdimensionen «Inhalt», «Textaufbau», «Sprachrichtigkeit» und «Sprachangemessenheit, Schreibstil und Kreativität» entwickelt. Diese vier Dimensionen entstanden in Anlehnung an das «Zürcher Textanalyseraster» von Nussbaumer und Sieber² sowie basierend auf dem Vorgehen von Becker-Mrotzek und Böttcher³. Das Zürcher Textanalyseraster erfasst die «sprachsystematische und orthographische Richtigkeit», die «funktionale Angemessenheit», die «ästhetische Angemessenheit» und die «inhaltliche Relevanz». Für die pädagogische Praxis erwies sich das Zürcher Analyseraster als zu detailliert. Zudem bietet es für die inhaltlichen Belange zu wenig Anhaltspunkte für die Beurteilung von Schülertexten⁴. Aus diesen Gründen schlagen Becker-Mrotzek und Böttcher einen Basiskatalog mit zwölf Kriterien vor, verteilt auf die fünf Basisdimensionen «Sprachrichtigkeit», «Sprachangemessenheit», «Inhalt», «Aufbau» und «Schreibprozess». Unter der Basisdimension «Schreibprozess» werden nach Becker-Mrotzek und Böttcher die folgenden beiden Kriterien zusammengefasst: «Planen/Überarbeiten» (Lässt der Text Pla-

² Nussbaumer, M. & Sieber, P. (1994). Texte analysieren mit dem Zürcher Textanalyseraster. In P. Sieber (Hrsg.), *Sprachfähigkeiten – besser als ihr Ruf und nötiger denn je! Ergebnisse aus einem Forschungsprojekt* (S. 141–186). Aarau: Sauerländer.

³ Becker-Mrotzek, M. & Böttcher, I. (2011). *Schreibkompetenzen entwickeln und beurteilen*. Berlin: Cornelsen

⁴ Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11*. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen. Münster: Waxmann.

nungs- und Überarbeitungsspuren erkennen?) und «Wagnis/Kreativität» (Lässt der Text ein besonderes sprachliches Wagnis erkennen? Ist er in besonderer Weise kreativ?).

Anpassungen wurden entsprechend bisheriger Erfahrungen mit der Korrektur von Texten und aufgrund testtheoretischer Gütekriterien vorgenommen. Zudem wurden die inhaltlichen Kriterien auf die konkrete Aufgabenstellung angepasst. Das Beurteilungsverfahren entspricht einem analytischen Vorgehen, bei dem verschiedene Aspekte eines Textes unabhängig voneinander nach verbal formulierten Abstufungen bewertet werden⁵.

3.2 Beurteilungskriterien

In Tabelle 1 sind die Kriterien zur Beurteilung der Texte bezüglich der Dimension «Inhalt: Auftragserfüllung und Aussagekraft» ersichtlich. In dieser Dimension wurden bei den drei Themen inhaltlich analoge Bereiche beurteilt. Bewertet wurde, ob die Ausführungen zum gewählten Thema passen, inwiefern auf die gestellten Fragen eingegangen wird, wie gehaltvoll die Aussagekraft des Textes ist und ob die Schülerinnen und Schüler beim Verfassen des Textes auch ein inhaltliches Wagnis eingehen.

Tabelle 1: Kriterien des Beurteilungsrasters zur Dimension «Inhalt: Auftragserfüllung und Aussagekraft»

| Kriterien | Abstufungen |
|---------------------------------------|--|
| 1.1 Ausführungen | 1 passen teilweise zum gestellten Thema 2 passen weitgehend zum gestellten Thema 3 passen gut zum gestellten Thema |
| 1.2 Fragen | 1 werden teilweise beantwortet 2 werden weitgehend beantwortet 3 werden ausführlich, detailliert beantwortet |
| 1.3 Aussagekraft / Gehalt | 1 teilweise gehaltvoll 2 weitgehend gehaltvoll 3 gehaltvoll |
| 1.4 inhaltliches Wagnis / Kreativität | 1 wagt wenig, wenig kreativ 2 wagt etwas, teilweise kreativ 3 wagt viel, kreativ 4 wagt sehr viel, sehr kreativ |

⁵ Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

In Tabelle 2 sind die Kriterien zur Dimension «Textaufbau und Textzusammenhang» dargestellt. Es wurde beurteilt, ob ein Text in eine sinnvolle äussere Gliederung (Einleitung, Hauptteil, Schluss) eingeteilt ist. Zusätzlich wurde bewertet, ob der Text in sich logisch zusammenhängt (Kohärenz) und ob die Sätze und Abschnitte sprachlich sinnvoll verbunden sind (Kohäsion).

Tabelle 2: Kriterien des Beurteilungsrasters zur Dimension «Textaufbau und Textzusammenhang»

| Kriterien | Abstufungen |
|---|---|
| 2.1 Textaufbau (Abschnitte – äussere Gliederung) | 1 zufällig, unüberlegt, ungegliedert oder inkonsequent 2 gegliedert (Einleitung, Hauptteil und Schluss sind ersichtlich) |
| 2.2 logischer Zusammenhang (innere Gliederung – Kohärenz) | 1 teilweise logisch ausgeführte Gedanken 2 meistens logisch ausgeführte Gedanken 3 logisch ausgeführte Gedanken |
| 2.3 sprachlicher Zusammenhang (innere Gliederung – Kohäsion) | 1 teilweise sinnvoll verbunden 2 meistens sinnvoll verbunden 3 sinnvoll verbunden, auch bei komplexeren Verbindungen |

Tabelle 3 umfasst die Kriterien zur Dimension «Sprachrichtigkeit». Dabei wurde beurteilt, ob ein Text in Bezug auf Rechtschreibung, Grammatik, Satzbau und Satzzeichen korrekt ist.

Tabelle 3: Kriterien des Beurteilungsrasters zur Dimension «Sprachrichtigkeit»

| Kriterien | Abstufungen |
|---|--|
| 3.1 Rechtschreibung | 1 ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei |
| 3.2 Grammatik (Genus, Kasus, Tempus, Modus) | 1 ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei |
| 3.3 Satzbau (Korrektheit, Vollständigkeit, inkl. Kohäsionsmittel) | 1 ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei |
| 3.4 Satzzeichen (Satzschlusszeichen, Kommas, Anführungszeichen etc.) | 1 ansatzweise beherrscht 2 teilweise beherrscht 3 weitgehend beherrscht 4 nahezu fehlerfrei |

Tabelle 4 beinhaltet die Kriterien der Dimension «Sprachangemessenheit». Bei dieser Dimension wurde bewertet, ob die Wortwahl angemessen und treffsicher ist und ob der Satzbau abwechslungsreich ist. Zudem wurde beurteilt, wie gewandt sich die Schülerinnen und Schüler ausdrücken (Sprachstil) sowie ob sie sprachlich etwas wagen.

Tabelle 4: Kriterien des Beurteilungsrasters zur Dimension «Sprachangemessenheit, Schreibstil und Kreativität»

| Kriterien | Abstufungen |
|--|--|
| 4.1 Satzbau | 1 teilweise abwechslungsreich 2 abwechslungsreich 3 sehr abwechslungsreich, vielseitig |
| 4.2 Wortwahl | 1 teilweise treffend, teilweise angemessen 2 treffend, angemessen 3 sehr treffsicher, angemessen, auch bei komplexeren Begriffen |
| 4.3 Schreibstil | 1 sprachlich wenig gewandt 2 sprachlich teilweise gewandt 3 sprachlich gewandt 4 sprachlich sehr gewandt, ausdrucksstark |
| 4.4 Sprachliches Wagnis – Kreativität und Ästhetik | 1 wagt wenig, wenig kreativ 2 wagt etwas, teilweise kreativ 3 wagt viel, kreativ 4 wagt sehr viel, sehr kreativ |

4 Qualität des Beurteilungsrasters

4.1 Analyse der Beurteilungskriterien

In Tabelle 5 ist die Qualität der Beurteilungskriterien der Dimension «Inhalt: Auftragserfüllung und Aussagekraft» dargestellt. Die Trennschärfen⁶ der vier Beurteilungskriterien liegen alle über 0.30 und können somit als ausreichend bis gut bewertet werden. Das Kriterium 1.1 ist eher einfach ausgefallen, das bedeutet, dass es von vielen Schülerinnen und Schülern gut erfüllt wird. 78 Prozent der Schülerinnen und Schüler haben in ihren Texten Ausführungen gemacht, welche gut zum gestellten Thema passen. Am schwersten fiel den Schülerinnen und Schüler das inhaltliche Wagnis beziehungsweise die inhaltliche Kreativität (Kriterium 1.4).

⁶ Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.

Tabelle 5: Qualität der Kriterien zur Dimension « Inhalt: Auftragserfüllung und Aussagekraft»

| Beurteilungskriterien | Abstufungen | Schwierigkeit | Trennschärfe | Infit (MNSQ) | T-Wert |
|---------------------------------------|-------------|---------------|--------------|--------------|--------|
| 1.1 Ausführungen | 1 | 3% | 0.35 | 1.10 | 3.4 |
| | 2 | 19% | | | |
| | 3 | 78% | | | |
| 1.2 Fragen | 1 | 9% | 0.44 | 1.08 | 3.9 |
| | 2 | 56% | | | |
| | 3 | 36% | | | |
| 1.3 Aussagekraft / Gehalt | 1 | 5% | 0.53 | 0.97 | -1.4 |
| | 2 | 52% | | | |
| | 3 | 43% | | | |
| 1.4 inhaltliches Wagnis / Kreativität | 1 | 4% | 0.54 | 0.98 | -0.8 |
| | 2 | 52% | | | |
| | 3 | 40% | | | |
| | 4 | 4% | | | |

Tabelle 6 beinhaltet die Darstellung der Qualität der Beurteilungskriterien der Dimension «Textaufbau und Textzusammenhang». Die Trennschärfe der beiden Kriterien zum logischen und sprachlichen Zusammenhang (Kriterien 2.2 und 2.3) ist als sehr gut zu bewerten. Das Kriterium zum Textaufbau (2.1) weist hingegen eine zu geringe Trennschärfe auf. Dies bedeutet, dass nur ein geringer Zusammenhang zwischen der Schreibkompetenz und der Strukturierung der Texte in Einleitung, Hauptteil und Schluss besteht. Schülerinnen und Schüler mit guten Schreibkompetenzen strukturieren ihre Texte somit nicht zwingend entsprechend dieser Dreiteilung. Generell gliederten nur 53 Prozent der Schülerinnen und Schüler ihren Text so, dass eine Einleitung, ein Hauptteil und ein Schluss ersichtlich waren. Die beiden Kriterien zum logischen und sprachlichen Zusammenhang passen gut zum Raschmodell. Zudem verteilen sich die Texte bei diesen beiden Kriterien relativ gleichmässig auf die beiden oberen Abstufungen (2 und 3), während die unterste Stufe (1) selten vertreten ist. Dies bedeutet, dass nur wenige Texte aus teilweise logischen Gedanken bestanden und die Sätze und Abschnitte in den Texten nur selten wenig sinnvoll miteinander verbunden waren.

Tabelle 6: Qualität der Kriterien zur Dimension «Textaufbau und Textzusammenhang»

| Beurteilungskriterien | Abstufungen | Schwierigkeit | Trennschärfe | Infit (MNSQ) | T-Wert |
|-------------------------------|-------------|---------------|--------------|--------------|--------|
| 2.1 Textaufbau | 1 | 47% | 0.25 | 1.17 | 13.8 |
| | 2 | 53% | | | |
| 2.2 logischer Zusammenhang | 1 | 3% | 0.56 | 0.92 | -4.6 |
| | 2 | 43% | | | |
| | 3 | 54% | | | |
| 2.3 sprachlicher Zusammenhang | 1 | 5% | 0.63 | 0.84 | -7.9 |
| | 2 | 68% | | | |
| | 3 | 27% | | | |

In Tabelle 7 ist die Qualität der Beurteilungskriterien der Dimension «Sprachrichtigkeit» dargestellt. Die Trennschärfe der vier Beurteilungskriterien (Rechtschreibung, Grammatik, Satzbau und Satzzeichen) ist gut. Beim Kriterium Satzzeichen verteilen sich die Texte relativ gleichmässig auf alle vier Abstufungen, während bei der Rechtschreibung, der Grammatik und beim Satzbau die unterste Stufe (1) eher seltener vertreten ist. Dieses Ergebnis bedeutet, dass wenige Schülerinnen und Schüler die Rechtschreibung, die Grammatik und den Satzbau nur ansatzweise beherrschen, während das korrekte Setzen von Satzzeichen von vergleichsweise etwas mehr Schülerinnen und Schüler nur ansatzweise beherrscht wird.

Tabelle 7: Qualität der Kriterien zur Dimension «Sprachrichtigkeit»

| Beurteilungskriterien | Abstufungen | Schwierigkeit | Trennschärfe | Infit (MNSQ) | T-Wert |
|-----------------------|-------------|---------------|--------------|--------------|--------|
| 3.1 Rechtschreibung | 1 | 8% | 0.56 | 1.13 | 6.0 |
| | 2 | 21% | | | |
| | 3 | 43% | | | |
| | 4 | 28% | | | |
| 3.2 Grammatik | 1 | 4% | 0.55 | 1.13 | 5.8 |
| | 2 | 13% | | | |
| | 3 | 36% | | | |
| | 4 | 47% | | | |
| 3.3 Satzbau | 1 | 4% | 0.61 | 1.01 | 0.6 |
| | 2 | 15% | | | |
| | 3 | 39% | | | |
| | 4 | 42% | | | |
| 3.4 Satzzeichen | 1 | 14% | 0.55 | 1.16 | 7.8 |
| | 2 | 34% | | | |
| | 3 | 36% | | | |
| | 4 | 16% | | | |

Tabelle 8 beinhaltet die Darstellung der Qualität der Beurteilungskriterien der Dimension «Sprachangemessenheit, Schreibstil und Ästhetik». Die Trennschärfe der vier Kriterien ist als gut bis sehr gut zu bewerten und die Kriterien passen gut zum Rasch-Modell (Infit). Bei den ersten beiden Kriterien zum Satzbau und zur Wortwahl sind die Texte hauptsächlich in der mittleren Stufe (2) eingeordnet, während die unterste Stufe (1) seltener vertreten ist. Dies bedeutet, dass die meisten Schülerinnen und Schüler einen abwechslungsreichen Satzbau verwenden und ihre Wortwahl treffend und angemessen ist. Sprachlich sehr sicher und sehr gewandt (Kriterium 4.3, Stufe 4) sind nur wenige Texte. Dasselbe gilt auch für das Kriterium 4.4 zum sprachlichen Wagnis sowie zur Kreativität und Ästhetik.

Tabelle 8: Qualität der Kriterien zur Dimension «Sprachangemessenheit, Schreibstil und Ästhetik»

| Beurteilungskriterien | Abstufungen | Schwierigkeit | Trennschärfe | Infit (MNSQ) | T-Wert |
|--|-------------|---------------|--------------|--------------|--------|
| 4.1 Satzbau | 1 | 7% | 0.51 | 0.96 | -1.6 |
| | 2 | 71% | | | |
| | 3 | 22% | | | |
| 4.2 Wortwahl | 1 | 5% | 0.55 | 0.90 | -4.2 |
| | 2 | 77% | | | |
| | 3 | 18% | | | |
| 4.3 Schreibstil | 1 | 6% | 0.79 | 0.72 | -15.5 |
| | 2 | 46% | | | |
| | 3 | 45% | | | |
| | 4 | 3% | | | |
| 4.4 Sprachliches Wagnis – Kreativität und Ästhetik | 1 | 2% | 0.55 | 0.94 | -2.9 |
| | 2 | 55% | | | |
| | 3 | 39% | | | |
| | 4 | 4% | | | |

4.2 Eindimensionalität des Beurteilungsrasters

Bei der Reliabilität erreicht das gesamte Beurteilungsraster ein sehr gutes Cronbach-Alpha von 0.81. Dies bedeutet, dass Schülerinnen und Schüler mit einer hohen Schreibkompetenz bei allen Kriterien – mit Ausnahme des Kriteriums zum Textaufbau – eine bessere Bewertung erhielten als Schülerinnen und Schüler mit einer geringeren Schreibkompetenz. Die Beurteilungskriterien sind somit konsistent angewendet worden und eignen sich gut, um zuverlässig zwischen guten und weniger guten Texten zu differenzieren, und um die Schreibkompetenz von Schülerinnen und Schülern zu bestimmen.

5 Beurteilungszuverlässigkeit

5.1 Zusammensetzung des Korrekturteams

Die Texte der Schülerinnen und Schüler wurden von sieben Korrektorinnen und Korrektoren (Raterinnen und Rater) bewertet. Rater A studierte Pädagogik sowie Deutsche Sprach- und Literaturwissenschaft. Rater B und C studierten Deutsche Sprach- und Literaturwissenschaft. Rater D studierte Germanistik, Romanistik und Philosophie und promovierte in Germanistik. Rater E ist ausgebildete Sekundarlehrperson phil. I mit mehrjähriger Erfahrung. Rater G studierte Germanistik und Filmwissenschaften. Rater F ist ausgebildeter Primarlehrer. Die Zuordnung der Texte zu den Raterinnen und Ratern erfolgte klassenweise nach dem Zufallsprinzip.

5.2 Verfahren der Mehrfachkorrekturen

Es besteht die Gefahr, dass sich bei verschiedenen Raterinnen und Ratern über die Zeit hinweg ein unterschiedliches mentales Modell einzelner Beurteilungskriterien entwickelt. Daher ist es wichtig in regelmässigen Abständen das gemeinsame Verständnis der Beurteilungskriterien zu überprüfen⁷. Dies wurde dadurch gewährleistet, indem täglich ca. zehn Texte von allen anwesenden Raterinnen und Ratern beurteilt wurden (Mehrfachkorrekturen). Das Ziel war dabei, eine möglichst hohe Übereinstimmung der Beurteilung der Texte zwischen den Raterinnen und Ratern zu erreichen. Insgesamt wurden mit diesem Verfahren rund 250 Texte von mehreren Raterinnen und Ratern beurteilt. Die Abweichungen in der Beurteilung wurden fortlaufend an die Raterinnen und Rater zurückgemeldet und mit ihnen besprochen.

5.3 Übereinstimmung in der Bewertung

In Tabelle 9 sind die Übereinstimmungen beziehungsweise die Abweichungen zwischen den Raterinnen und Ratern bei den Mehrfachkorrekturen dargestellt. Die Prozentzahl der Übereinstimmung kommt über einen Vergleich von allen sieben Raterinnen und Ratern untereinander zustande (21 mögliche Kombinationen). Die Spalte «Übereinstimmung» beinhaltet den Anteil der vollständigen Übereinstimmung zweier Raterinnen oder Ratern, die darauf folgenden Spalten enthalten dementsprechend den Anteil an Abweichungen um einen, zwei oder drei Punkte pro Kriterium.

Wie Tabelle 9 zeigt, wurde die höchste prozentuale Übereinstimmung beim Kriterium Textaufbau (2.1) erreicht. Dies bedeutet, dass sich die Raterinnen und Rater in 84 Prozent der Fälle einig waren, ob ein Text ungegliedert ist, oder ob die Gliederung in Einleitung, Hauptteil und Schluss ersichtlich ist. Die geringste prozentuale Übereinstimmung mit 48 Prozent liegt bei der Beurteilung der Korrektheit des Satzbaus (Kriterium 3.3) vor. Bei diesem Kriterium waren sich die Raterinnen und Rater nur in etwa der Hälfte aller Texte einig. Die angestrebte möglichst hohe Übereinstimmung zwischen den Personen des Korrekturteams wurde insgesamt gesehen erreicht. Bei den beiden Kriterien «Ausführungen» (1.1) und «Textaufbau» (2.1) wurde die angestrebte Übereinstimmung sehr gut erreicht. Bei den Kriterien «Satzbau: Korrektheit» (3.3) und «Aussagekraft / Gehalt» (1.3) ist die Übereinstimmung hingegen noch verbesserungsfähig.

⁷ Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64–86.

Tabelle 9: Übereinstimmung und Abweichungen in der Bewertung

| Beurteilungskriterien | Übereinstimmung | Abweichung 1 Punkt | Abweichung 2 Punkte | Abweichung 3 Punkte |
|---|-----------------|-----------------------|------------------------|------------------------|
| 1.1 Ausführungen | 74% | 25% | 1% | – |
| 1.2 Fragen | 57% | 40% | 3% | – |
| 1.3 Aussagekraft / Gehalt | 49% | 48% | 3% | – |
| 1.4 Inhaltliches Wagnis / Kreativität | 56% | 41% | 4% | 0% |
| 2.1 Textaufbau | 84% | 16% | 0% | – |
| 2.2 logischer Zusammen- hang | 56% | 42% | 2% | – |
| 2.3 sprachlicher Zusammen- hang | 61% | 38% | 1% | – |
| 3.1 Rechtschreibung | 55% | 42% | 3% | 0% |
| 3.2 Grammatik | 56% | 38% | 5% | 0% |
| 3.3 Satzbau: Korrektheit | 48% | 44% | 7% | 0% |
| 3.4 Satzzeichen | 53% | 42% | 5% | 0% |
| 4.1 Satzbau: Abwechslung | 58% | 40% | 2% | – |
| 4.2 Wortwahl | 67% | 32% | 1% | – |
| 4.3 Schreibstil | 56% | 43% | 1% | 0% |
| 4.4 Sprachliches Wagnis – Kreativität und Ästhetik | 57% | 41% | 3% | 0% |

5.4 Strenge der Bewertungen

Mit der Anwendung der Item-Response-Theorie⁸ ist es zusätzlich möglich, die Korrektur-Strenge der beurteilenden Personen ins Testmodell einzubeziehen und bei der Berechnung der Ergebnisse zu berücksichtigen. Die Strenge der Raterinnen und Rater wird dabei auf einer logarithmischen Skala abgebildet und als «Logit» bezeichnet. Ein positiver Logit entspricht einer hohen Strenge, ein negativer Logit einer geringen Strenge. Wie Tabelle 10 zeigt, variiert die Strenge der Raterin-

⁸ Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.

nen und Rater zwischen Rater E mit einem Logit von -0.292 und Rater B mit einem Logit von 0.468. Rater E hat somit am wenigsten streng korrigiert, während Rater B die Texte der Schülerinnen und Schüler am strengsten beurteilt hat. Die Rater-Strengung wird bei der Berechnung der Ergebnisse für die Schülerinnen und Schüler mitberücksichtigt.

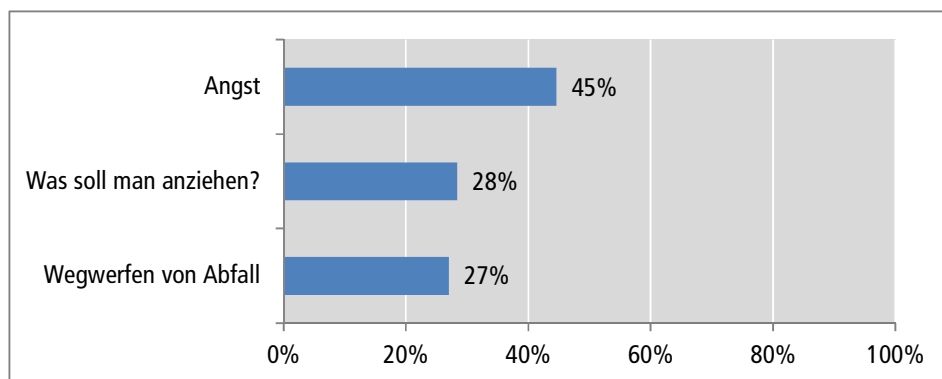
Tabelle 10: Strengung und Modellkonformität der Korrekturen

| Raterin / Rater | Strengung (Logit) | Schätzfehler | Infit (MNSQ) | T-Wert |
|-----------------|-------------------|--------------|--------------|--------|
| Rater A | 0.145 | 0.012 | 1.08 | 1.4 |
| Rater B | 0.468 | 0.011 | 1.03 | 0.7 |
| Rater C | -0.006 | 0.012 | 0.92 | -1.7 |
| Rater D | -0.255 | 0.013 | 1.27 | 4.4 |
| Rater E | -0.292 | 0.013 | 1.17 | 3.0 |
| Rater F | -0.163 | 0.012 | 1.07 | 1.5 |
| Rater G | 0.103 | 0.030 | 1.40 | 6.5 |

5.5 Auswahl und Schwierigkeit der Themen

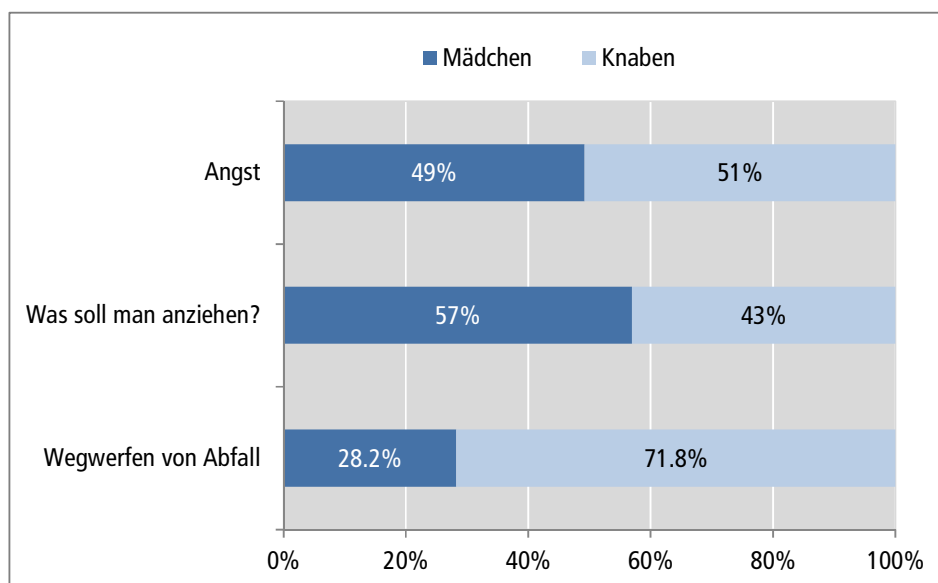
Abbildung 3 zeigt, welches Thema wie häufig gewählt wurde. Das Thema «Angst» wurde mit 45 Prozent am häufigsten gewählt. Die beiden Themen «Was soll man anziehen?» und «Bussen für das Wegwerfen von Abfall» wurden fast gleichhäufig von 28 Prozent respektive 27 Prozent der Schülerinnen und Schülern gewählt.

Abbildung 3: Auswahl der Themen



Insgesamt haben 2376 Knaben (54%) und 2001 Mädchen (46%) einen Text verfasst. In der Abbildung 4 ist dargestellt, wie sich die beiden Geschlechter bei den drei Themen jeweils verteilen. Das Thema «Angst» haben in etwa gleich viele Knaben wie Mädchen gewählt. Das Thema «Was soll man anziehen?» war bei den Mädchen beliebter, während das Thema «Bussen für das Wegwerfen von Abfall» öfter von Knaben gewählt wurde.

Abbildung 4: Auswahl der Themen nach Geschlecht



Mit Hilfe der Raschskalierung ist es möglich, die Schwierigkeit der drei Themen zu schätzen und auf einer logarithmischen Skala als «Logit» abzubilden – analog zur Berechnung der Strenge der Raterinnen und Rater. Je positiver der Logit ausfällt, desto höher ist die Schwierigkeit eines Themas ausgeprägt (vgl. Tabelle 11). Die drei Themen unterscheiden sich in der Schwierigkeit nur sehr gering. Das Thema «Was soll man anziehen?» ist etwas schwieriger ausgefallen als die beiden anderen Themen. Dies bedeutet jedoch nicht zwingend, dass dieser Schreibauftrag von der Thematik her schwieriger ist, da auch die Möglichkeit besteht, dass dieses Thema in der Tendenz eher von schwächeren Schülerinnen und Schülern gewählt wurde. Die Wahl des Themas wird aufgrund der geringen Unterschiede in der Schwierigkeit in den weiteren Analysen beziehungsweise bei der Berechnung der Ergebnisse der Schülerinnen und Schüler nicht berücksichtigt.

Tabelle 11: Schwierigkeit und Modellkonformität der beiden Themen

| Thema | Schwierigkeit (Logit) | Schätzfehler | MNSQ | T-Wert |
|-------------------------------------|-----------------------|--------------|------|--------|
| Angst | -0.010 | 0.008 | 1.03 | 1.0 |
| Was soll man anziehen? | 0.119 | 0.008 | 0.97 | -0.6 |
| Bussen für das Wegwerfen von Abfall | -0.110 | 0.011 | 1.02 | 0.4 |

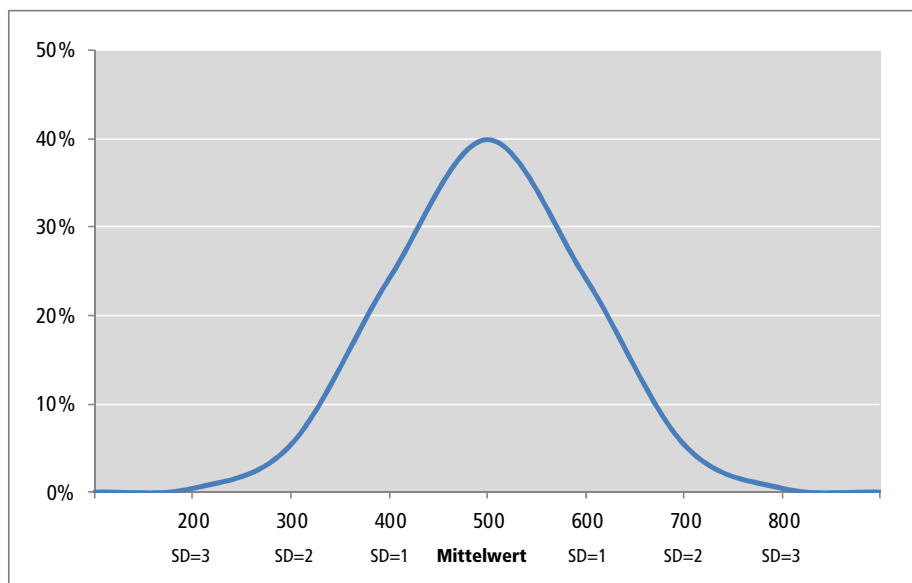
6 Testergebnisse

Dass derselbe Text trotz vorgegebener, standardisierter Beurteilungskriterien, sorgfältiger Schulung der Raterinnen und Rater sowie regelmässiger Mehrfachkorrekturen von mehreren Personen immer gleich beurteilt wird, ist aufgrund des Interpretationsspielraums bei offen gestellten Aufgaben nicht zu erwarten. Bei der Beurteilung eines Textes bestimmen vier Faktoren das Testergebnis: a) Die Schreibkompetenz der Schülerin oder des Schülers, b) die Schwierigkeit der Beurteilungskriterien, c) die Korrektur-Strenge der Raterinnen und Rater und d) das Thema. Die Fähigkeiten der Schülerinnen und Schüler sowie die Korrektur-Strenge der Raterinnen und Rater werden bei der Berechnung der Ergebnisse berücksichtigt.

6.1 Verteilung der Ergebnisse auf der Stellwerkskala

Um die Ergebnisse innerhalb des Moduls «Texte schreiben» vergleichbar zu machen, wurden die Fähigkeiten der Schülerinnen und Schüler in eine standardisierte Normalverteilung transformiert, welche analog zur Stellwerk-Skala, einen Mittelwert von 500 und eine Standardabweichung (SD) von 100 Punkten aufweist (siehe Abbildung 5). Diese Skala hat die Eigenschaft, dass rund 68 Prozent der Ergebnisse zwischen 400 und 600 Punkten liegen, rund 95 Prozent zwischen 300 und 700 Punkten und nahezu alle Ergebnisse zwischen 200 und 800 Punkten. Die Punktzahl auf der Stellwerk-Skala zeigt den Schülerinnen und Schülern, wie gut sie innerhalb der Vergleichsgruppe der 4377 Schülerinnen und Schüler aus der 8. Klasse im Kanton Aargau abschnitten.

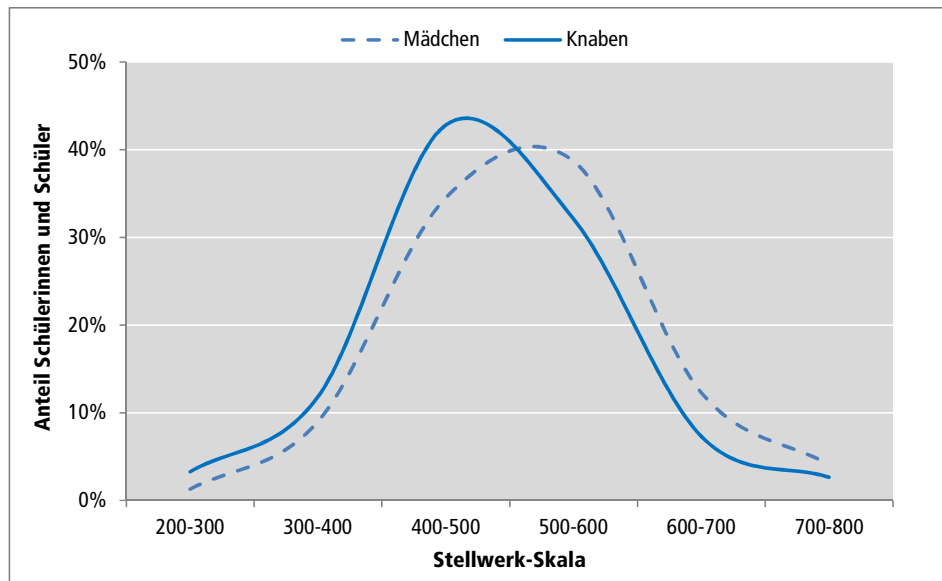
Abbildung 5: Verteilung der Punkte auf der Stellwerkskala



6.2 Verteilung der Ergebnisse nach Geschlecht

Abbildung 6 zeigt die Verteilung der Fähigkeiten der Schülerinnen und Schüler im Modul «Texte schreiben» als Punktzahl auf der Stellwerkskala, getrennt nach Geschlecht. Die Mädchen erreichen im Durchschnitt 515 Punkte auf der Stellwerkskala, während die durchschnittliche Punktzahl der Knaben bei 486 liegt. Diese Differenz von 29 Punkten ist statistisch signifikant⁹ ($p < .001$), jedoch von geringer Bedeutsamkeit (Effektstärke $d = 0.30$).

Abbildung 6: Fähigkeiten der Schülerinnen und Schüler nach Geschlecht



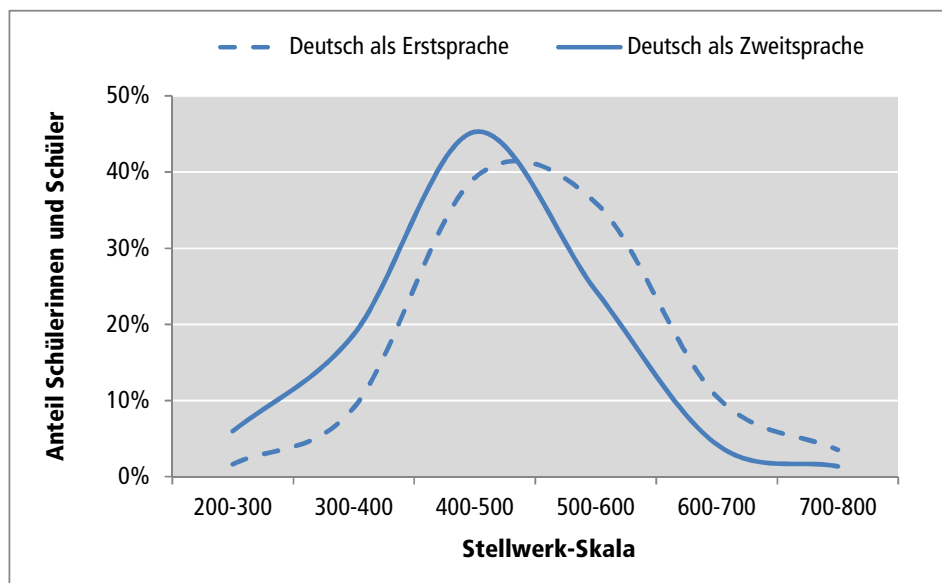
6.3 Ergebnisse nach Erstsprache

Abbildung 7 zeigt die Verteilung der Punktzahl auf der Stellwerkskala nach der Erstsprache der Schülerinnen und Schüler. Der Mittelwert der Schülerinnen und Schüler mit Deutsch als Erstsprache liegt bei 506 Punkten, jener der Schülerinnen und Schüler mit Deutsch als Zweitsprache bei 457 Punkten. Der Unterschied von 49 Punkten ist statistisch signifikant und als eher gross zu beurteilen (Effektstärke $d = 0.50$).

Die Unterschiede nach Erstsprache sind allerdings nicht in allen Schultypen gleich gross. In der Bezirksschule beträgt der Vorsprung der Schülerinnen und Schüler mit Deutsch als Erstsprache 38 Punkte ($d = 0.46$), in der Sekundarschule 14 Punkte ($d = 0.18$) und in der Realschule 21 Punkte ($d = 0.25$). Der Unterschied in der Schreibkompetenz zwischen Schülerinnen und Schüler mit Deutsch als Erstsprache und Deutsch als Zweitsprache ist somit in der Bezirksschule bedeutsamer als in der Sekundar- und Realschule.

⁹ Die statistischen Begriffe in diesem Kapitel werden im Glossar im Anhang dieses Berichtes erläutert.

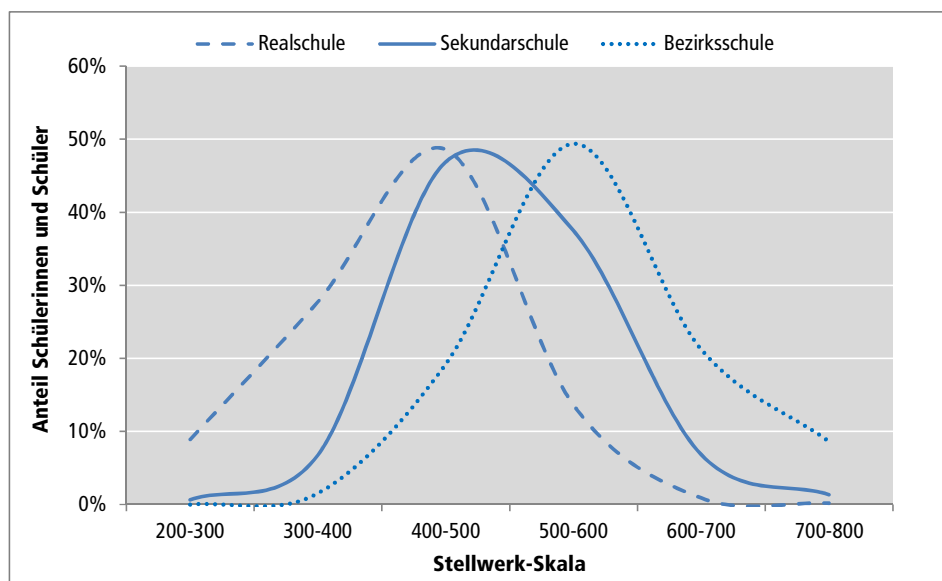
Abbildung 7: Fähigkeiten der Schülerinnen und Schüler nach Erstsprache



6.4 Ergebnisse nach Schultyp

Abbildung 8 zeigt die Ergebnisse nach Schultyp. In der Bezirksschule beträgt der Mittelwert 567 Punkte, in der Sekundarschule 499 Punkte und in der Realschule 419 Punkte. Aus Abbildung 8 geht hervor, dass sich die Leistungen in den verschiedenen Schultypen überlappen, insbesondere im mittleren Leistungsbereich. Der Unterschied von 80 Punkten zwischen der Real- und der Sekundarschule ist signifikant und sehr gross ($p < .001$, Effektstärke $d = 1.00$). Etwas kleiner ist der Unterschied mit 68 Punkten (Effektstärke $d = 0.86$) zwischen der Sekundar- und der Bezirksschule. Auch dieser Unterschied ist statistisch signifikant ($p < .001$).

Abbildung 8: Fähigkeiten der Schülerinnen und Schüler nach Schultyp



7 Beispieltexte zu den einzelnen Kompetenzniveaus

Um die Kompetenzen der einzelnen Schüler standardisiert beschreiben zu können, wurden die Punktzahlen auf der Stellwerkskala zu Kompetenzniveaus zusammengefasst (siehe Tabelle 12). Die Schreibkompetenzen, die eine Schülerin oder ein Schüler innerhalb eines bestimmten Kompetenzniveaus aufweist, wurden zum einen bezüglich der drei Teilgebiete «Inhalt», «Textaufbau» und «Sprache» beschrieben und zum anderen mit Beispieltexten illustriert. Die Beschreibung der Schreibkompetenzen auf den verschiedenen Niveaus bezieht sich jeweils auf einen durchschnittlichen Text des jeweiligen Niveaus. Bei konkreten Einzelleistungen von Schülerinnen und Schülern kann es vorkommen, dass die Kompetenzen innerhalb einer Gesamtpunktzahl je nach Teilbereich variieren. Beispielsweise könnte ein Text des Intervalls 501 bis 600 Punkte im Teilbereich «Inhalt» tiefer (z.B. im Kompetenzniveau von 401 bis 500 Punkten) und dafür im Teilbereich «Sprache» etwas höher (z.B. im Kompetenzniveau von 601 bis 700 Punkten) liegen. Innerhalb eines bestimmten Kompetenzniveaus besteht somit eine recht grosse Bandbreite an Texten, die sich innerhalb der einzelnen Teilbereiche unterscheiden können. In der folgenden Zusammenstellung wird jedes Kompetenzniveau mit einem Beispieltexten veranschaulicht.

Tabelle 12: Kompetenzbeschreibungen nach Punkteintervallen

| Punkteintervall | Kompetenzbeschreibungen | | |
|--------------------|--|--|--|
| | Inhalt | Textaufbau | Sprache |
| 200 bis 300 Punkte | Die Texte sind teilweise gehaltvoll. Die in der Aufgabenstellung verlangten Inhalte sind erkennbar und die Ausführungen passen teilweise zum gestellten Thema. Die Fragen der Aufgabenstellung werden teilweise beantwortet. | Der Textaufbau ist zufällig und ungegliedert. Die Sätze und Abschnitte bestehen teilweise aus logisch ausgeführten Gedanken und sind teilweise sinnvoll miteinander verbunden. | Die Rechtschreibung und die Zeichensetzung werden ansatzweise beachtet. Grammatikalische und syntaktische Strukturen werden ansatzweise beherrscht. |
| 301 bis 400 Punkte | Die Texte erfüllen die Aufgabenstellung recht gut, sind recht gehaltvoll und inhaltlich teilweise kreativ. Die in der Aufgabenstellung verlangten Fragen werden recht gut beantwortet. | Die Texte sind eher zufällig gegliedert, bestehen aber aus meistens logisch ausgeführten Gedanken. Die Sätze und Abschnitte sind oftmals sinnvoll miteinander verbunden. | Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung werden teilweise beachtet. Die Wortwahl ist meistens angemessen und der Satzbau meistens abwechslungsreich. Die sprachliche Ausdrucksweise ist teilweise gewandt und teilweise kreativ. |
| 401 bis 500 Punkte | Die Texte erfüllen die Aufgabenstellung weitgehend, sind weitgehend gehaltvoll und inhaltlich recht kreativ. Die in der Aufgabenstellung verlangten Fragen werden weitgehend beantwortet. | Die Texte sind gegliedert und bestehen meistens aus logisch ausgeführten Gedanken. Zumeist werden auch die Textelemente sinnvoll miteinander verbunden. | Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung werden weitgehend beherrscht. Die Wortwahl ist treffend und angemessen, der Satzbau abwechslungsreich. Der sprachliche Ausdruck ist recht gewandt und kreativ. |
| 501 bis 600 Punkte | Die Texte erfüllen die Aufgabenstellung gut, sind gehaltvoll und inhaltlich kreativ. Die in der Aufgabenstellung verlangten Fragen werden recht ausführlich beantwortet. | Die Texte sind sinnvoll gegliedert; eine Dreiteilung in Einleitung, Hauptteil und Schluss ist erkennbar. Die Texte bestehen fast immer aus logisch ausgeführten Gedanken und die Textelemente werden sinnvoll miteinander verbunden. | Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung sind fast fehlerfrei. Die Wortwahl ist treffsicher, der Satzbau abwechslungsreich und die sprachliche Ausdrucksweise gewandt und kreativ. |
| 601 bis 700 Punkte | Die Texte erfüllen die Aufgabenstellung sehr gut, sind gehaltvoll und inhaltlich kreativ. Die Ausführungen passen gut zum gestellten Thema und die in der Aufgabenstellung verlangten Fragen werden ausführlich und detailliert beantwortet. | Die Texte sind zusätzlich zur Grundgliederung in weitere inhaltliche Sinnschritte gegliedert, ein roter Faden ist erkennbar. Die Gedankenführung ist logisch und die Textelemente werden auch bei komplexeren Formulierungen sinnvoll miteinander verbunden. | Die Rechtschreibung, die Grammatik, der Satzbau und auch die Zeichensetzung sind nahezu fehlerfrei. Die Wortwahl ist auch bei komplexeren Begriffen treffsicher und angemessen. Die Texte enthalten abwechslungsreiche Formulierungen, sind sprachlich gewandt und kreativ. |
| 701 bis 800 Punkte | Die Texte gehen ausführlich, detailliert und elaboriert auf die Aufgabenstellung ein. Die Fragen werden einleuchtend und ausführlich beantwortet. Inhaltlich enthalten die Texte unerwartete Ausführungen und sind ausgesprochen kreativ und gehaltvoll. | Die Texte haben einen roten Faden, folgen einem logischen, klaren Aufbau und sind äusserlich und innerlich sinnvoll gegliedert. Die Textelemente werden auch bei komplexeren Verbindungen logisch und sinnvoll miteinander verbunden. | Die Texte sind auch bei komplexeren Formulierungen nahezu fehlerfrei. Die sprachliche Ausdrucksweise ist sehr gewandt und ausdrucksstark. Die Texte zeichnen sich durch ansprechende, abwechslungsreiche und kreative Sprachstrukturen sowie unerwartete Formulierungen aus. |

Textbeispiel 1: 200 bis 300 Punkte

Die Texte sind teilweise gehaltvoll. Die in der Aufgabenstellung verlangten Inhalte sind erkennbar und die Ausführungen passen teilweise zum gestellten Thema. Die Fragen der Aufgabenstellung werden teilweise beantwortet. Der Textaufbau ist zufällig und ungegliedert. Die Sätze und Abschnitte bestehen teilweise aus logisch ausgeführten Gedanken und sind teilweise sinnvoll miteinander verbunden. Die Rechtschreibung und die Zeichensetzung werden ansatzweise beachtet. Grammatikalische und syntaktische Strukturen werden ansatzweise beherrscht.

Bussen für abfälle im öffentlichen Raum sollte bestraft werden es ist nicht gut für uns es ist ^{nicht} mal ~~1~~ für die Umwelt. Beobachte sie mal die Umwelt es liegt schon viel abfall an den Bahnhöfen, es könnte viel schöner auf der Erde aussehen. Es werden immer mehr abfälle im öffentlichen Raum weggeschmissen. Meine Erfahrungen zeigen das Leute viel zu faul sind ihr Müll so richtig weg zu schmeissen und deshalb hab ich Erfahrung mit diesen Menschen. meine Meinung ist es sollte kleine Bussen geben.

Textbeispiel 2: 301 bis 400 Punkte

Die Texte erfüllen die Aufgabenstellung recht gut, sind recht gehaltvoll und inhaltlich teilweise kreativ. Die in der Aufgabenstellung verlangten Fragen werden recht gut beantwortet. Die Texte sind eher zufällig gegliedert, bestehen aber aus meistens logisch ausgeführten Gedanken. Die Sätze und Abschnitte sind oftmals sinnvoll miteinander verbunden. Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung werden teilweise beachtet. Die Wortwahl ist meistens angemessen und der Satzbau meistens abwechslungsreich. Die sprachliche Ausdrucksweise ist teilweise gewandt und teilweise kreativ.

Ich habe Angst um mein Leben

Ich Mr. Harris war ein Millioner Schwere Mann in Los Angeles sehr bekannt für mein Drogenhandel und meine kriminellen Taten bekannt. Ich wohne in einer großen Villa mit 12 asiatischen Frauen. Ich tue alles mit anderen Gangs zutun aber wenig mit der Polizei. Die Polizei will sich nicht mit uns anlegen. ist auch besser so sonst gebe es jeden Tag Tote und das will die Polizei und auch nicht. aber dennoch gibt es Tote wenn sich Gangs nicht freitragen. Zurzeit bin ich im Krankenhaus so schickte sie jetzt das einander den Text für mich schreiben m. u. s. es begann am 13.5.2004 an ich bekam eine SMS von einer anderen Gang The Last. Ich habe sie fertig erget und jetzt wollen sie sich rechnen. Ich habe keine Angst ich habe sie dann schon fertig gemacht und heute werde ich sie auch fertig machen. 2 Stunden später für ein nigel neuer langer ging in das eingangsfahr. Es waren 4 Männer maskiert und bewaffnet. Sie zogen die Waffen und erschossen die eiskalt.

zu diesem Zeitpunkt war ich unbewaffnet Sie kamen zum mir und verhafteten mich sie schleppten mich in den Keller bindeten meine Hände an die empfinden. Der einte fand Schlagstöcke. Mit diesem verhafteten sie mich doch ich bin ich mir einarm und Grippe ich habe Angst um mein Leben.

Textbeispiel 3: 401 bis 500 Punkte

Die Texte erfüllen die Aufgabenstellung weitgehend, sind weitgehend gehaltvoll und inhaltlich recht kreativ. Die in der Aufgabenstellung verlangten Fragen werden weitgehend beantwortet. Die Texte sind gegliedert und bestehen meistens aus logisch ausgeführten Gedanken. Zumeist werden auch die Textelemente sinnvoll miteinander verbunden. Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung werden weitgehend beherrscht. Die Wortwahl ist treffend und angemessen, der Satzbau abwechslungsreich. Der sprachliche Ausdruck ist recht gewandt und kreativ.

Was soll man anziehen?

Viele Jugendliche tragen gerne Markenklieder, die Vorbilder tragen oder Stars.

Bei meinem Outfit ist eigentlich alles wichtig. Doch vor allem muss die Grösse passen und der Stoff muss bequem sein. Bei mir spielt das Geld nicht eine grosse Rolle. Das Material soll nicht schon nach einer Woche kaputt sein. Ich muss mich wohl fühlen.

In der Schule habe ich mehrere Beobachtungen gemacht. Alle Jugendlichen sind verschieden angezogen. Es gibt Mädchen/Knaben die mit Markenklieder angezogen sind und die mit ~~kein~~ normal angezogen sind. Es spielt keine Rolle ob du mit Markenklieder oder Markenschuhe hast, aber wenn man anfängt die Jugendlichen auszulachen dann finde ich es gemein und Respektlos. Jeder hat sein Style!

Man kann für mich alles anziehen ob du fester oder dünner bist man muss sich wohl fühlen. Ich finde einfach man sollte Respekt haben, weil wenn man ausgelacht wird fühlt sich man nicht mehr wohl. Dann denkt man „ich sollte vielleicht mein Style ändern“. Doch würde jeder genau gleich ange-

zogen sein oder den genau gleichen ^{Look} Style, dann würde es langweilig werden!

Textbeispiel 4: 501 bis 600 Punkte

Die Texte erfüllen die Aufgabenstellung gut, sind gehaltvoll und inhaltlich kreativ. Die in der Aufgabenstellung verlangten Fragen werden recht ausführlich beantwortet. Die Texte sind sinnvoll gegliedert; eine Dreiteilung in Einleitung, Hauptteil und Schluss ist erkennbar. Die Texte bestehen fast immer aus logisch ausgeführten Gedanken und die Textelemente werden sinnvoll miteinander verbunden. Die Rechtschreibung, die Grammatik, der Satzbau und die Zeichensetzung sind fast fehlerfrei. Die Wortwahl ist treffsicher, der Satzbau abwechslungsreich und die sprachliche Ausdrucksweise gewandt und kreativ.

Ich war schon den ganzen Tag unterwegs, weil ich am Vortag bei Jeremy übernachtet habe. Als wir^{nun} um die Blöcke umrundeten rief Jeremy Lukas an und fragte ob er rauskommt. Worauf Lukas seine Sachen anzog und gleich rauskam. Nun zu dritt machten wir uns auf den Weg und als ob der kalte Herbstwind nicht schon genug ziehen würde fing es auch noch wie ~~tau~~ tausend Wasserfälle an zu regnen. Ich rannte als erster in Schutz worauf ~~wir~~ mir die Beiden folgten. Wir liefen unter den Verandas durch bis hin zur Migro wo wir ~~trud~~ uns in der Tiefgarage nieder ließen. Kurz darauf juckte Lukas auf und lief weg wir fragten uns was den los war. Ich hörte ein dumpfes Geräusch. „Es kommt von der Treppe!“ sagte Jeremy und schaute nach. Da kam ~~aus~~ Lukas aus den Tür mit einem Sach Erde. „Was das soll“ sagte er uns nicht er verteilte die Erde nur fröhlich durch die Garage. Jeremy gefiel seine Idee und hatte Holzhohle, dass alles vor der Migro gestapelt war. Auch er fing an die Wände mit Kohle zu bemalen. Kurz darauf kriegte ich ein

unwohltes Gefühl im Bauch, doch ich machte trotzdem mit. Wir breiteten zur Krönung WC Papier in der ganzen Garage aus. Ich war so vertieft in diese Szene das ich nicht bemerkte das zwei junge ~~Männer~~ ^{Männer} in Sportkleidung hinter uns standen. Jeremy schrie rennt! Und erst als ich diesen Schrei hörte wusste ich, ~~ich~~ wir sind geliefert. Ich hatte solch eine Angst, ich rannte und sah neben mir Lukas rennen. Wir rannten weg von diesem Ort und verschwanden hinter einer Hausecke. Lukas lehnte sich an und schnaubte laut. Ich aufgepumpt mit Adrenalin stand da total geschockt als die Polizei einfuhr. Erneut rannten wir weg hinter uns die Polizei und die Beiden Männer, deren Motorräder wie es sich herausstellte wir mit WC Papier eingewickelt haben. Ich rannte über alles drüber, rempelte Leute und kletterte über Zäune. Einer der Männer sprang und packte sich im Flug Jeremy. Ich dachte nur jetzt ist es gelaufen, mir ging durch den Kopf anzuhalten doch es war keine Zeit.

Das Adrenalin strömte immer noch durch meinen Körper und als Lukas und ich immer noch von der Polizei verfolgt durch eine enge Gasse rannten stiegen die Cops aus und blickten uns nur hinterher. Ich hatte Angst und lief zu einer Kollegin der ich dann alles schilderte. Sie sagte mir ich soll keine Angst haben und abwarten. ~~ka~~

In dieser Sekunde klingelte das Handy von mir, es war die Polizei.

Wir horchten ohne ein Wort von uns zu geben, da sagte eine raue Stimme:

"Komm* sofort zur Migro wenn nicht gibt es Geldstrafe und Gemeinnützige Arbeit!"

Wir liefen also ohne Wahl zur ~~na~~ Migro und liessen uns ordentlich was anhören.

Am meisten ~~an~~ Angst hatte ich vor meiner Mutter die auftauchte. Sie schimpfte mich an und starrte mich nur noch an.

Ich sank den Kopf und putzte zur - Strafe mit den 'zwei die ganze Garage.

Textbeispiel 5: 601 bis 700 Punkte

Die Texte erfüllen die Aufgabenstellung sehr gut, sind gehaltvoll und inhaltlich kreativ. Die Ausführungen passen gut zum gestellten Thema und die in der Aufgabenstellung verlangten Fragen werden ausführlich und detailliert beantwortet. Die Texte sind zusätzlich zur Grundgliederung in weitere inhaltliche Sinnschritte gegliedert, ein roter Faden ist erkennbar. Die Gedankenführung ist logisch und die Textelemente werden auch bei komplexeren Formulierungen sinnvoll miteinander verbunden. Die Rechtschreibung, die Grammatik, der Satzbau und auch die Zeichensetzung sind nahezu fehlerfrei. Die Wortwahl ist auch bei komplexeren Begriffen treffsicher und angemessen. Die Texte enthalten abwechslungsreiche Formulierungen, sind sprachlich gewandt und kreativ.

~~Jeder hat ein~~ Jeder sollte selbst entscheiden dürfen, was man selbst anziehen will. Manche mögen es lieber schwarz, mit viel Blumen oder einen eher altmodischen Look.

Mir persönlich ist es bei einem Outfit wichtig, dass es bequem ist und es dem Tag angepasst ist. Zum Beispiel würde ich nie ein Abendkleid in die Schule oder Jogginghosen und einen Sport-BH zu einem Fest anziehen. Mir ist es auch wichtig, dass ich mich wohl fühle, in dem was ich trage, denn darunter leidet vielleicht auch das Selbstbewusstsein. Beim Einkauf achte ich auf gute Qualität und nicht unbedingt auf die Marke, wobei ich beim Schmuck doch eher zu teureren Sachen greife, da man Schmuck immer tragen kann und er eigentlich nie zu klein wird, denn deshalb denke ich, dass man beim Schmuck mal tiefer in den Geldbeutel greifen kann. Was ich aber persönlich schade finde und auch immer wieder auf der Straße, im Restaurant oder in der Stadt sehe, ist dass einige Mädchen, zu wenig anziehen was schnell billig überkommen kann oder sie zu häufig hohe Schuhe anziehen und in diesen versuchen zu laufen, der Versuch aber meistens missglückt. Viele Mädchen tragen aber auch immer mehr Designersachen die viele zu teuer sind und man sicher für weniger Geld, gleichwertige oder vielleicht sogar bessere Sachen dafür kaufen kann. Man sollte von meiner Sicht aus her eher keine Imitate von Designern kaufen bzw. tragen, weil man den Fake eigentlich immer gut erkennen kann und wenn man sich dann schon das Original leisten kann, wird es häufiger als Fake bezeichnet weil es fast jeder zweite hat.

Was man sicher oft anziehen soll oder wenigstens im Schrank haben muss, sind viele ~~Star~~ Accessoires wie Gürtel, Schals, Hüte, etc. denn mit diesen kleinen Dingen kann man einen normalen Jeans und ein weisses Basic top gleich aufwerten und es sieht gleich viel besser aus.

Viele Leute würden meine Meinung sicher als falsch bezeichnen aber jeder hat einen anderen Stil und ein anderes Modebewusstsein.

Textbeispiel 6: 701 bis 800 Punkte

Die Texte gehen ausführlich, detailliert und elaboriert auf die Aufgabenstellung ein. Die Fragen werden einleuchtend und ausführlich beantwortet. Inhaltlich enthalten die Texte unerwartete Ausführungen und sind ausgesprochen kreativ und gehaltvoll. Die Texte haben einen roten Faden, folgen einem logischen, klaren Aufbau und sind äusserlich und innerlich sinnvoll gegliedert. Die Textelemente werden auch bei komplexeren Verbindungen logisch und sinnvoll miteinander verbunden. Die Texte sind auch bei komplexeren Formulierungen nahezu fehlerfrei. Die sprachliche Ausdrucksweise ist sehr gewandt und ausdrucksstark. Die Texte zeichnen sich durch ansprechende, abwechslungsreiche und kreative Sprachstrukturen sowie unerwartete Formulierungen aus.

Angst

Ich bin ein ganz normales 14-jähriges Mädchen, habe aber grosse Angst vor der Dunkelheit, wie auch der Nacht. Manchmal hatte ich sogar Halluzinationen, die nicht mehr aufhören wollten. Ich würde dir auch gerne von einem Erlebnis vor einer Woche erzählen.

Es war an einem Donnerstagabend, als ich mich in mein weiches Bett fallen liess. Ich packte mich mit der Decke bis zu den Schultern zu, da es in meinem Zimmer immer ziemlich kühl war. Als ich meine Augen schloss, begannen Sie. Die Halluzinationen. Ich hörte plötzlich Schritte, die von der obersten Treppe zu kommen schienen. Sie wurden immer lauter und es fühlte sich an als würden meine Ohren gleich Explodieren. Mir wurde plötzlich auch ganz heiss. Der Schweiß rollte über mein kaltes weisses Gesicht. Diese Schritte machten mich noch verrückt! Ich bin nicht verrückt!, sagte ich mir, um mich zu beruhigen und in der Hoffnung, die Schritte würden so verschwinden. Doch das taten Sie nicht, nein, ganz im Gegenteil! Sie kamen näher. Ich versuchte meine Angst zu unterdrücken und wegzuhören. Doch es ging nicht, Ich wusste nicht wieso, aber die Angst stieg in mir auf wie ein lodendes Feuer, welches immer grösser wird und immer mehr Dinge, mit sich

in seinen Schlund zieht. Ich glaube so ist es mit der Angst genau gleich. Es ist als würde sich die Angst mit unserer eigenen ~~At~~ Angst Energie tanken. Genau dann, als ich über das alles nachdachte, bemerkte ich, dass keine Schritte zu hören waren. Ich legte mich wieder beruhigt und nichts ahnend was nachher geschehen würde, hin. Ich schloss meine Augen und dachte darüber nach, wie mein Leben wohl in 6 Jahren aussehen würde. Ob ich immer noch im gleichen Bett einschlafen, wie auch aufwachen würde? Noch im selben Moment, als ich dies dachte, begann ich wieder die Schritte zu hören. Sie kamen aber nicht näher sonder schienen auf der gleichen Stelle zu gehen, was mich auf irgendeine Art und Weise beruhigt. Wahrscheinlich deshalb, weil sie nicht in meiner Nähe waren. Ich dachte mir, dass ich nicht wieder verzweifeln solle, wie beim ersten mal. Ich nahm meinen ganzen Mut, der noch übrig war, zusammen und stieg aus dem Bett. Zu meiner Sicherheit nahm ich noch einen Regenschirm mit mir. Ich ging ganz langsam die Treppe hoch, einen Schritt nach dem anderen, vorsichtig tapsend waren meine Schritte. Ich lugte um die Ecke, wo ich die Schritte vermutete. Doch da war nirgends jemand zu sehen. „Flavia, was machst du da?“, fragte mich eine allzu bekannte Stimme, die mich aber trotzdem zusammen zucken liess. Es war meine Mutter, welche mit verschränkten Armen hinter mir stand und mich vorwurfsvoll anblickte. „Ich habe Schritte gehört“, antwortete ich ihr wahrheitsgemäss. Sie schaute mich mit einem undefinierbaren Gesichtsausdruck an. Wir standen noch eine Weile so da, als sie sich plötzlich in Bewegung setzte und meinte: „Komm wir setzen uns Liebes, dann kannst du mir alles erzählen.“ Gesagt, getan. Am Ende meiner Erzählung nahm sie mich in den Arm und flüsterte mir ins Ohr: „Das sind nur Halluzinationen, welche dein Unter-

Unterbewusstsein erschafft. Davor musst du aber nicht Angst haben. Sie sind nicht echt, das musst du dir einfach merken!" Ich nicke beruhigt. „Könnte ich aber trotzdem bei dir schlafen?“, fragte ich meine Mutter. Sie nickte und so schliessen wir schon bald Arm in Arm ein, ohne wieder die Schritte zu hören.

Das, diese Halluzinationen hatte ich vor einer Woche. Am Anfang hatte ich eine heiden Angst aber trotzdem habe ich den Mut genommen und bin nach oben gegangen. Mit meiner Mutter darüber zu reden hat mir sehr geholfen, da ich das Gefühl hatte sie würde mich verstehen. Vielleicht war ihr ja genau dasselbe passiert? Aber das weiss ich nicht. Ich habe mir gemerkt, was meine Mutter mir gesagt hat: Halluzinationen sind nicht real, sondern nur Einbildung, welche unter Unterbewusstsein erschafft!

8 Fazit

Ende April 2014 wurde im Kanton Aargau das Stellwerk-Modul «Texte schreiben» als Ergänzung zu den computerbasierten Stellwerk-Tests durchgeführt. Mit diesem Modul werden die produktiven Fähigkeiten im Fach Deutsch, das heisst die Schreibkompetenzen, erfasst. Insgesamt nahmen 4377 Schülerinnen und Schüler der 8. Klasse teil.

Die Schülerinnen und Schüler konnten zwischen drei verschiedenen Themen wählen: 1) «Angst», 2) «Was soll man anziehen?» und 3) «Bussen für das Wegwerfen von Abfall». Bei allen drei Themen dienten den Schülerinnen und Schülern drei Fragen als Leitfaden für das Verfassen der Texte. Ebenfalls zum Schreibauftrag gehörten Angaben darüber, wie die Texte bewertet wurden.

Wie die Analysen zeigen, eignet sich das eingesetzte Beurteilungsraster gut zur Bewertung der Fähigkeiten im Bereich «Texte schreiben». Die Beurteilungskriterien sind, bis auf eine Ausnahme, alle ausreichend trennscharf. Das heisst, dass die eingesetzten Beurteilungskriterien wesentlich dazu beitragen, die Schreibkompetenzen der Schülerinnen und Schüler zu bestimmen, und zwar so, dass gute Schülerinnen und Schüler eine hohe Punktzahl erreichen und schwache Schülerinnen und Schüler eine tiefe Punktzahl.

Die Beurteilungsübereinstimmung der korrigierenden Personen (Rater) wurde mit verschiedenen Verfahren überprüft. Die Übereinstimmung zwischen den Raterinnen und Ratern im Rahmen der Doppelkorrekturen fällt bei einzelnen Items etwas geringer aus als erwartet. Allerdings liegt diese Abweichung zum grössten Teil im Rahmen von plus/minus einem Punkt pro Kriterium. Zudem hat die bisherige Forschung gezeigt, dass es trotz intensiver Schulung und Betreuung der Raterinnen und Rater schwierig ist, eine hohe Übereinstimmung zu erzielen¹⁰. Mit Hilfe der Item-Response-Theorie konnten die leicht unterschiedlichen Beurteilungsmassstäbe der Raterinnen und Rater bei der Berechnung der Ergebnisse der Schülerinnen und Schüler korrigiert werden, sodass eine faire Beurteilung gewährleistet ist.

Es zeigt sich im Rahmen des Moduls «Texte schreiben» dennoch, dass mit einem standardisierten Beurteilungsraster relativ einheitlich beurteilt werden kann. Die Beurteilungen können zudem aus testtheoretischer Perspektive als zuverlässig bezeichnet werden. Mithilfe der Item-Response-Theorie konnten die in der Strenge etwas variierenden Beurteilungsmassstäbe der Raterinnen und Rater bei der Berechnung der Ergebnisse der Schülerinnen und Schüler korrigiert werden, sodass eine faire Beurteilung möglich wurde.

Die Ergebnisrückmeldung auf der Stellwerk-Skala (Mittelwert = 500 Punkte, Standardabweichung = 100 Punkte) darf nicht darüber hinwegtäuschen, dass die Ergebnisse unabhängig von den anderen Testergebnissen der Stellwerk-Tests zu interpretieren sind. Der Mittelwert und die Standardabweichung beziehen sich ausschliesslich auf die 4377 beteiligten Schülerinnen und Schüler der 8. Klasse im Kanton Aargau. Die Mädchen erreichen im Durchschnitt signifikant bessere Ergebnisse als die Knaben. Die Darstellung der Ergebnisse nach Schultypen zeigt zudem, wie sinnvoll eine schultyp-unabhängige Beurteilung sein kann.

Abschliessend kann gesagt werden, dass die Erfassung der produktiven Kompetenzen im Fachbereich Deutsch aus einer testtheoretischen Perspektive als zuverlässig betrachtet werden darf. Wenn standardisierte Beurteilungskriterien sowie eine ausführliche Einarbeitungsphase der Rate-

¹⁰ O'Sullivan, B. & Rignall, M. (2007). Assessing the value of bias analysis feedback to raters for the IELTS Writing Modul. In L. Taylor & P. Falvey (Eds.), *IELTS collected papers: Research in speaking and writing assessment* (pp. 446–478). Cambridge, UK: Cambridge University Press.

rinnen und Rater mit einer klaren Definition des gemeinsamen Beurteilungsmasstabs eingehalten werden, kann die Erfassung von produktiven Kompetenzen mit der Erfassung von reproduktiven Kompetenzen mittels Leistungstests mithalten.

9 Glossar: Statistische Kennwerte

Effektgrösse – Punkte-Differenzen werden auf statistische Signifikanz geprüft. Die Signifikanz eines Ergebnisses sagt jedoch nichts aus über dessen Relevanz. Zur Interpretation von statistisch signifikanten Unterschieden wird die Effektgrösse «d» berechnet, indem die Differenz der Mittelwerte durch die Standardabweichungen dividiert wird. Punkte-Differenzen werden so standardisiert und vergleichbar. Eine Effektgrösse von $d = 0.2$ weist auf einen schwachen Effekt hin, eine Effektgrösse von $d = 0.5$ auf einen mittleren Effekt und eine Effektgrösse von $d = 0.8$ auf einen starken Effekt. Eine Punktedifferenz mit einem schwachen Effekt ist für die Praxis wenig bedeutsam, während Punkteunterschiede mit einem mittleren oder starken Effekt als bedeutsam interpretiert werden dürfen.

Infit – Der Infit (Weighted Mean Square, MNSQ) zeigt, wie gut die einzelnen Beurteilungskriterien zum Rasch-Modell passen. Er zeigt für jedes Kriterium, wie viele unerwartete Bewertungen unter der Annahme des Rasch-Modells beobachtet wurden. Der Infit hat ein Erwartungswert von 1.0. Weichen die Infit-Werte statistisch signifikant von 1.0 ab, dann passt das Beurteilungskriterium nicht zum Rasch-Modell. Ein zu hoher Infit-Wert weist darauf hin, dass die Trennschärfe des Beurteilungskriteriums zu niedrig ist. Ein guter Infit-Wert sollte in der Regel nicht kleiner als 0.7 und nicht grösser als 1.3 sein¹¹. Der T-Wert zeigt, ob die Infit-Werte statistisch signifikant vom erwarteten Wert von 1.0 abweichen. Ein Wert grösser als 1.96 beziehungsweise kleiner als -1.96 weist auf eine statistisch signifikante Abweichung hin. Die Bewertung eines Beurteilungskriteriums sollte sich allerdings nicht in erster Linie nach dem T-Wert richten, denn dieser hängt unter anderem auch von der Stichprobengrösse ab, welche beim vorliegenden Modul «Texte schreiben» mit 4377 Schülerinnen und Schüler sehr hoch ausfällt.

Rasch-Modell – Die Ergebnisse der Schülerinnen und Schüler werden mithilfe der Item-Response-Theorie berechnet. Mit der Anwendung dieses statistischen Modells (Rasch-Modell) ist es zusätzlich möglich, die Korrektur-Strenge der beurteilenden Personen sowie die Schwierigkeit der Themen ins Testmodell einzubeziehen und bei der Berechnung der Ergebnisse zu berücksichtigen. Ein solches Vorgehen wird auch als «Multi-Facetten-Modell» bezeichnet¹².

Reliabilität – Für die Reliabilität beziehungsweise die Messgenauigkeit eines Tests wird der Koeffizient «Cronbach-Alpha» berechnet. Das Cronbach-Alpha zeigt, wie stark die Beurteilungskriterien zusammenhängen. Werte über 0.70 werden als ausreichend, Werte über 0.80 als gut bezeichnet.

Schwierigkeit – Die Schwierigkeit eines Beurteilungskriteriums kann durch die Lösungshäufigkeit beziehungsweise den Anteil an Schülerinnen und Schülern, welche die einzelnen Abstufungen erreicht haben, dargestellt werden. Eine Lösungshäufigkeit zwischen 75 und 100 Prozent deutet auf ein (sehr) einfaches Kriterium hin, eine Lösungshäufigkeit von 50 bis 75 Prozent auf ein eher einfaches Kriterium, eine Lösungshäufigkeit zwischen 25 und 50 Prozent auf ein eher schwieriges Kriterium und eine Lösungshäufigkeit von weniger als 25 Prozent auf ein (sehr) schwieriges Kriterium.

¹¹ Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.

¹² Linacre, J. M. 1994. *Many-Facet Rasch Measurement*. Chicago, IL: MESA Press (original work published in 1989)

Statistische Signifikanz – Punkte-Differenzen werden auf statistische Signifikanz geprüft. Eine Differenz ist dann statistisch signifikant, wenn sie durch ein statistisches Testverfahren überprüft und, unter einer im Voraus festgelegten Irrtumswahrscheinlichkeit, beispielsweise 0.1%, für gültig befunden wird.

Trennschärfe – Unter der Trennschärfe wird die Korrelation eines Beurteilungskriteriums mit der Gesamtpunktzahl im Modul «Texte schreiben» verstanden. Der Trennschärfekoeffizient zeigt bei einem Test, inwiefern ein Bewertungskriterium Schülerinnen und Schüler mit hoher Punktzahl von Schülerinnen und Schülern mit niedriger Punktzahl trennt. Ein hoher Trennschärfekoeffizient (0.30 bis 1.00) bedeutet, dass Schülerinnen und Schüler mit einer hohen Punktzahl eine hohe Bewertung des jeweiligen Beurteilungskriteriums erreichten und solche mit einer niedrigen Punktzahl eine niedrige Bewertung. Ein niedriger Trennschärfekoeffizient (um 0) bedeutet, dass Schülerinnen und Schüler mit hohen und niedrigen Punktzahlen bei einem Kriterium gleich häufig eine hohe oder niedrige Bewertung erhielten. Ein negativer Trennschärfekoeffizient bedeutet, dass Schülerinnen und Schüler mit einer hohen Punktzahl eine niedrige Bewertung erhielten und solche mit einer niedrigen Punktzahl eine hohe Bewertung. Dementsprechend sollte der Trennschärfekoeffizient nicht kleiner als 0.30 sein. Angewendet auf die Beurteilung von Texten zeigt die Trennschärfe, wie gut die Punktzahl bei einem Kriterium mit der Gesamtpunktzahl übereinstimmt.